Maschinelles Lernen

InfoLab Saar
Universität des Saarlandes

Saarbrücken

Franz Walgenbach
Lukas Wachter

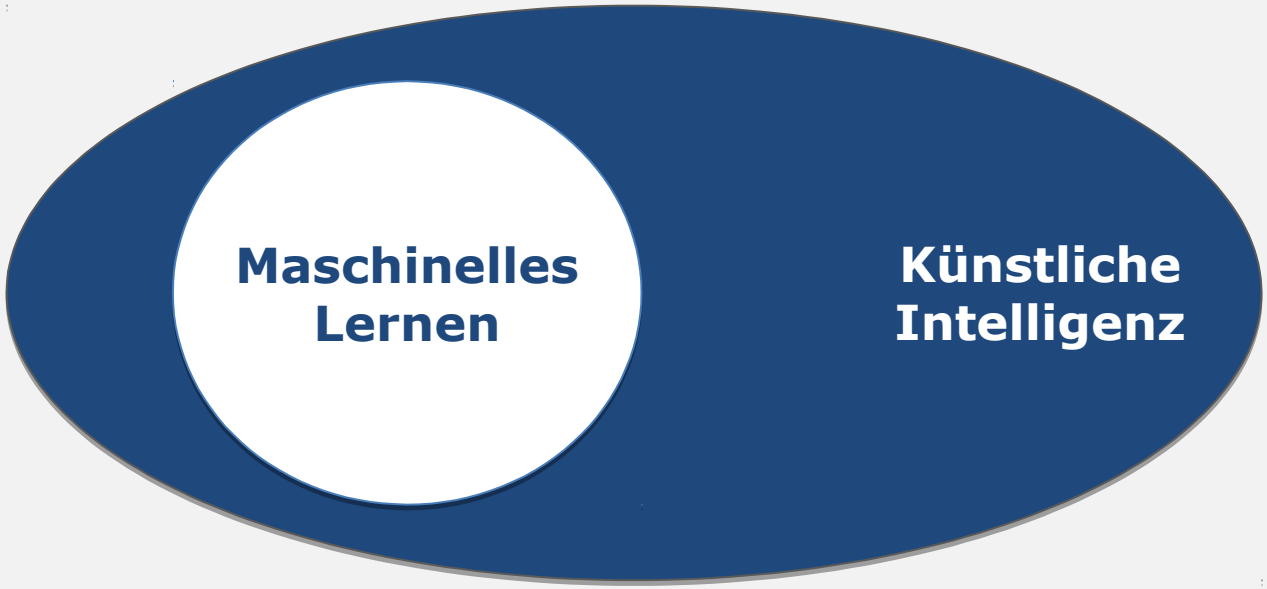
Self-Driving

ALPHAGO
AlphaGo
Lee Sedol

Health Care
Doctor
Hospital
Pharmacist
Nurse
Dentist
First Aid
Surgeon
Emergency
MEDICAL

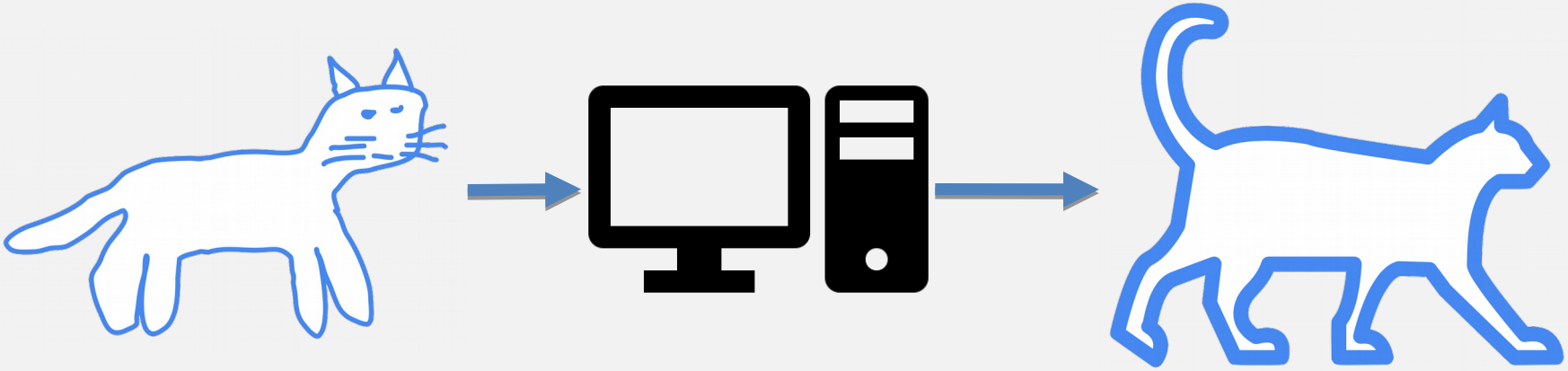
Künstliche Intelligenz beeinflusst mehr und mehr den Lebensalltag

Teilgebiet „Maschinelles Lernen“



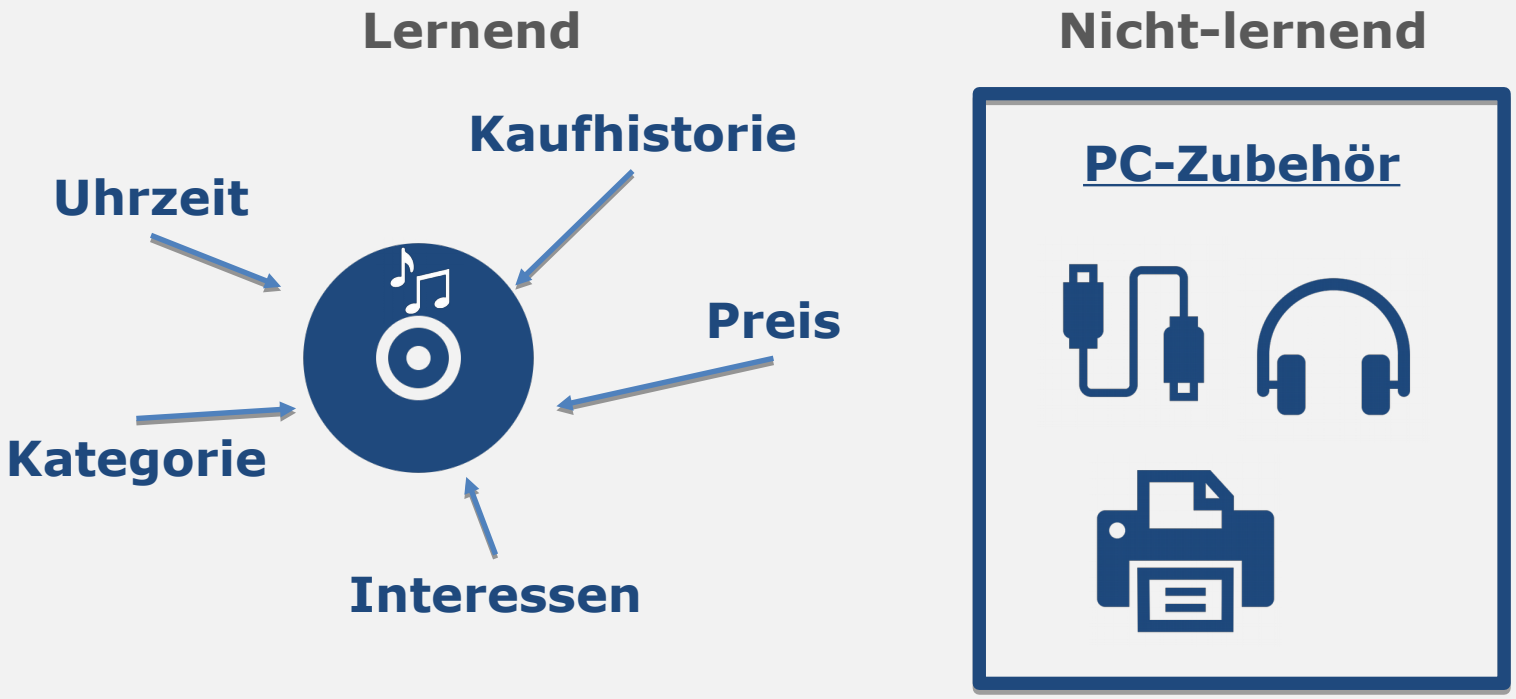
Was ist „Maschinelles Lernen“?

Google's Autodraw

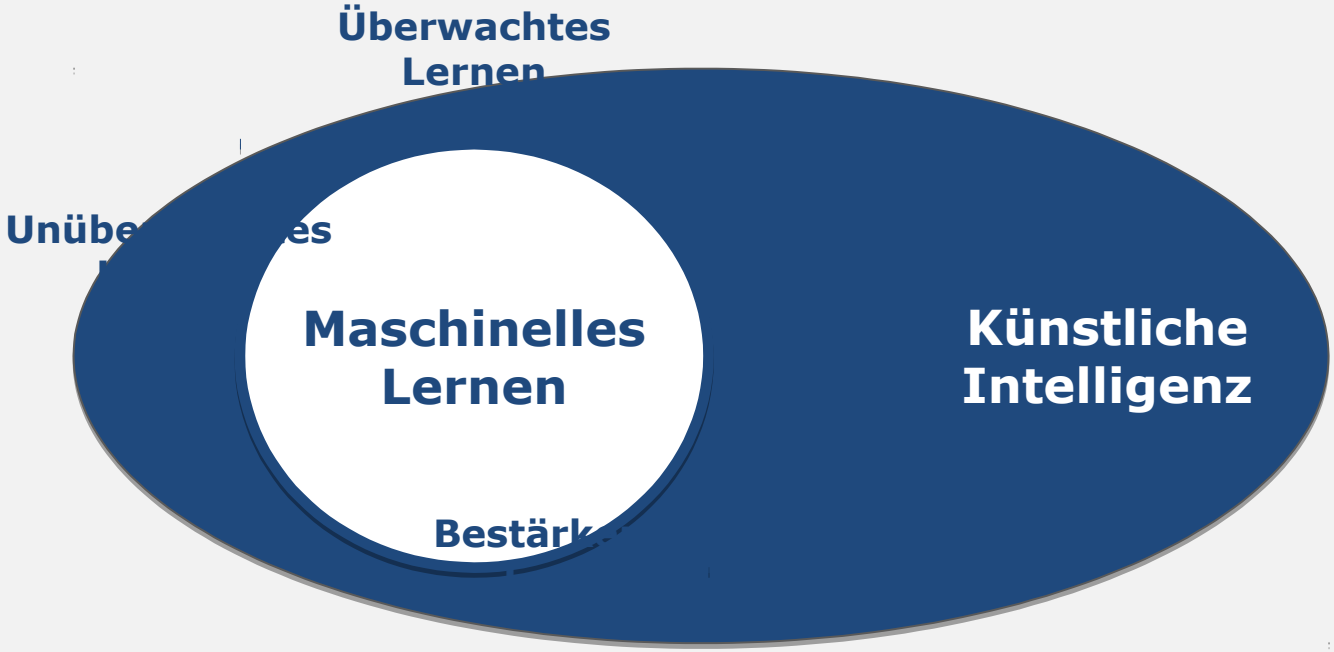


„Andere Kunden kauften auch …“

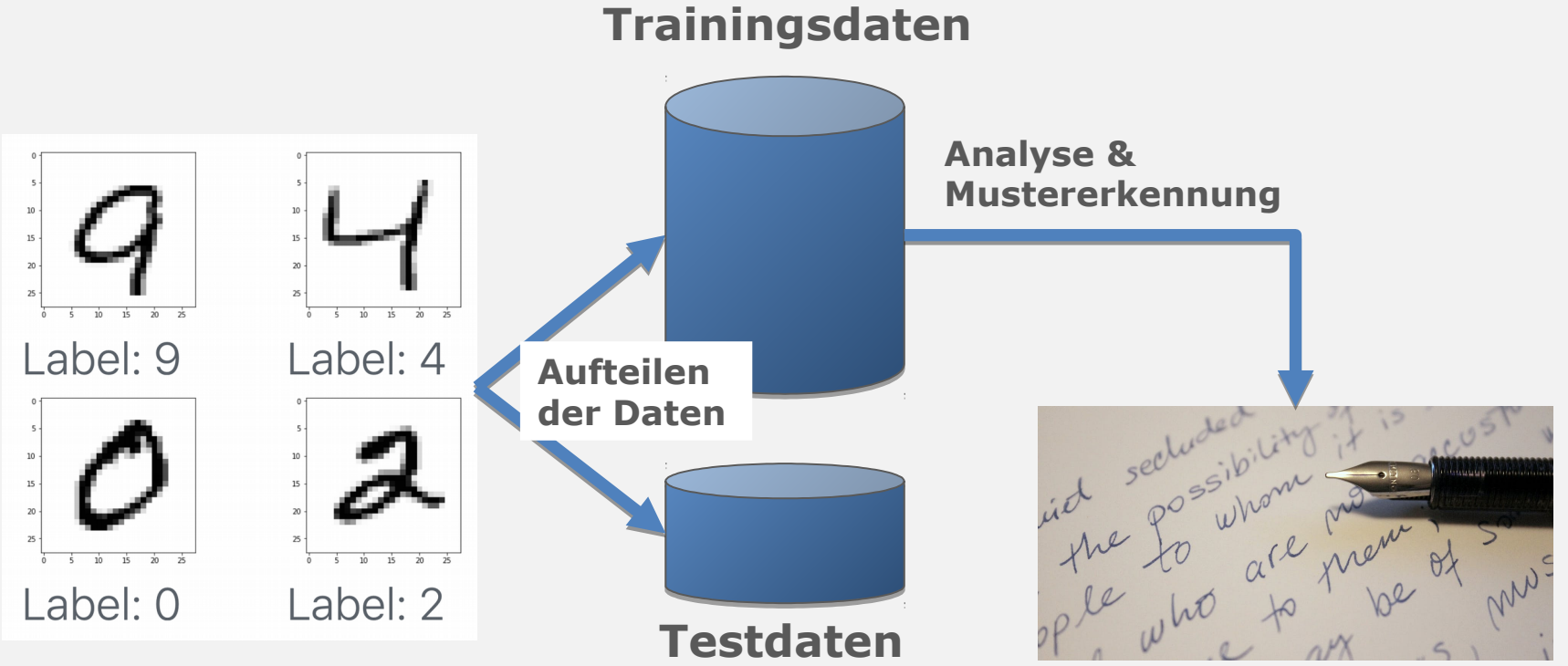
Situation: Computer bei Onlineshop gekauft, Preis: 899€



3 Typen maschinellen Lernens:



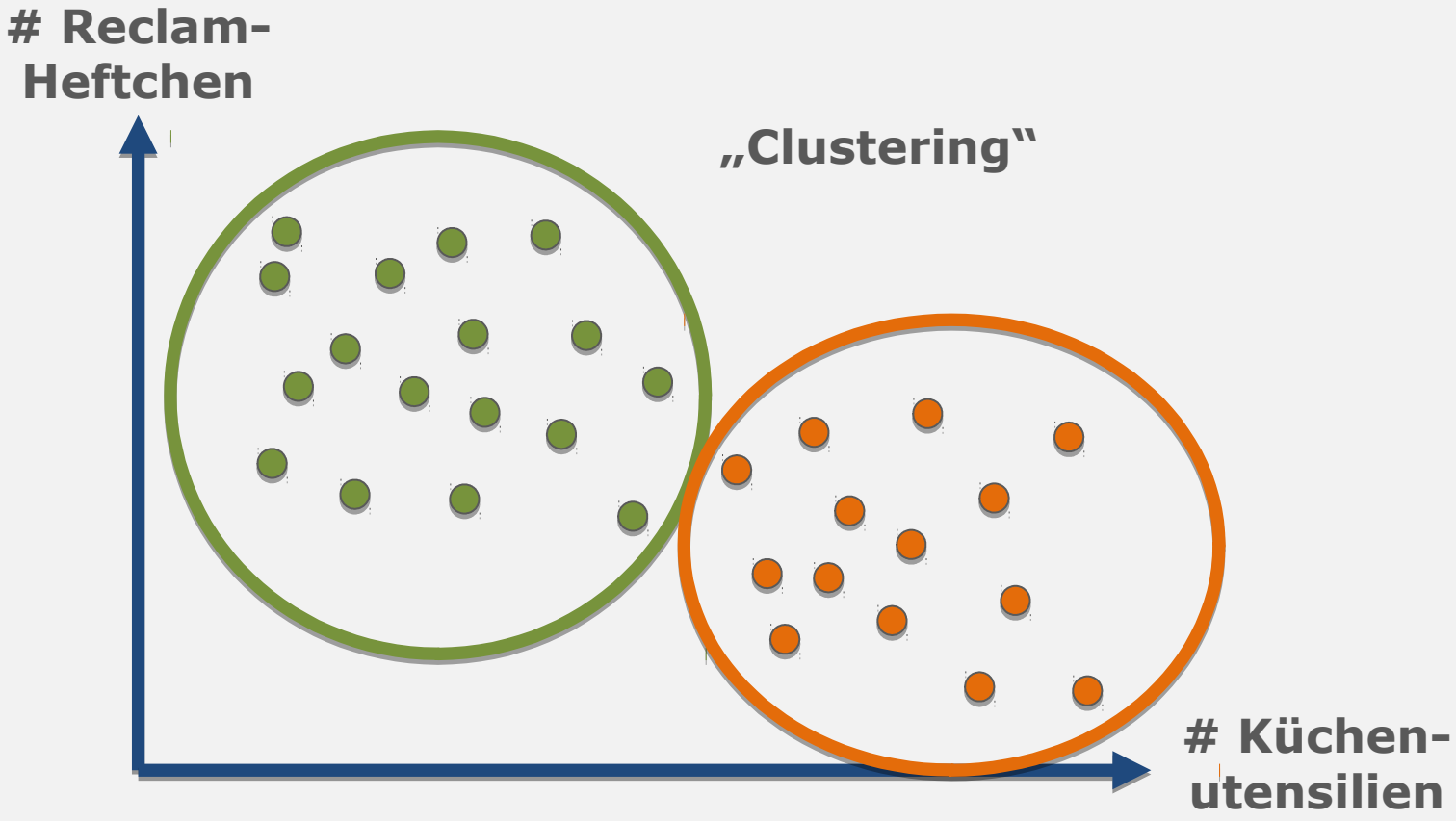
Überwachtes Lernen



Daten mit „Labels“

Erkennen neuer Daten

Unüberwachtes Lernen



Bestärkendes Lernen

Spielzüge analysieren und „besten Zug" auswählen

Sieg = Belohnung

Niederlage = Bestrafung

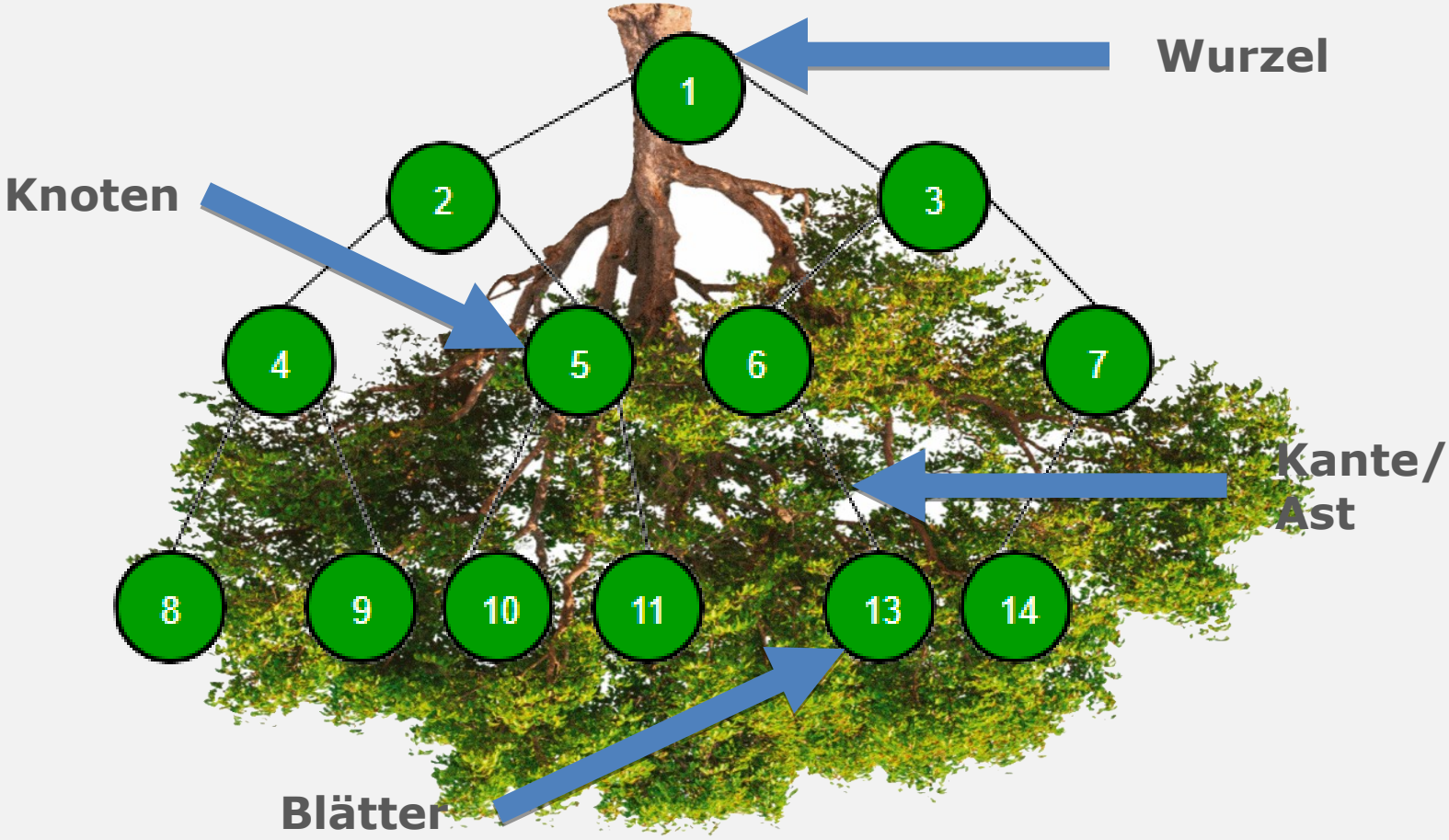
https://github.com/AmreshVenugopal/tic_tac_toe

Entspricht am ehesten dem menschlichen Lernprozess

W[illegible]s?

Entschei[illegible]ecision Tree)

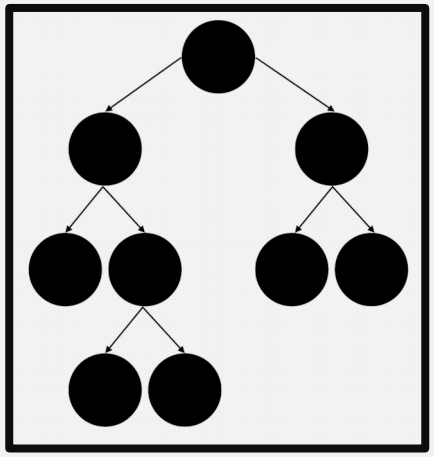
Bäume in der Informatik

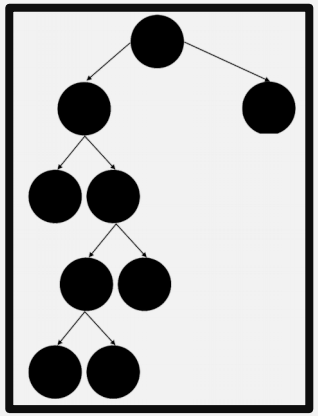


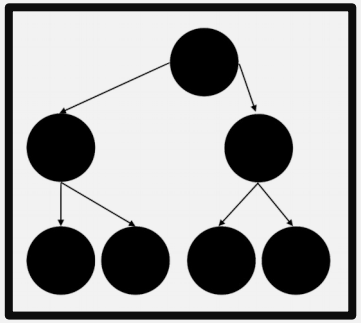
Was macht ein ENTSCHEIDUNGSbaum?

Beispiel: Hunde und Katzen

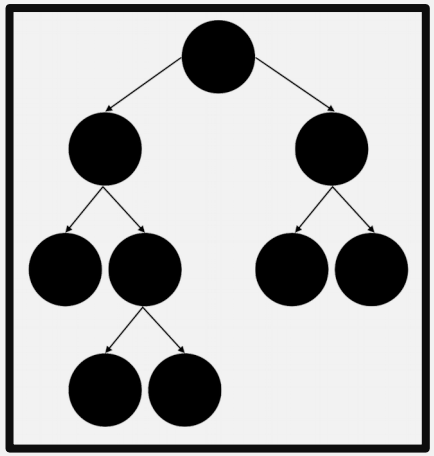
Decision Trees sind nicht eindeutig!







Baum mit der besten Genauigkeit auswählen



Ab jetzt führendes Beispiel:

Wettervorhersage

→ Überwachtes Lernen

Führendes Beispiel:

Trainingsdaten

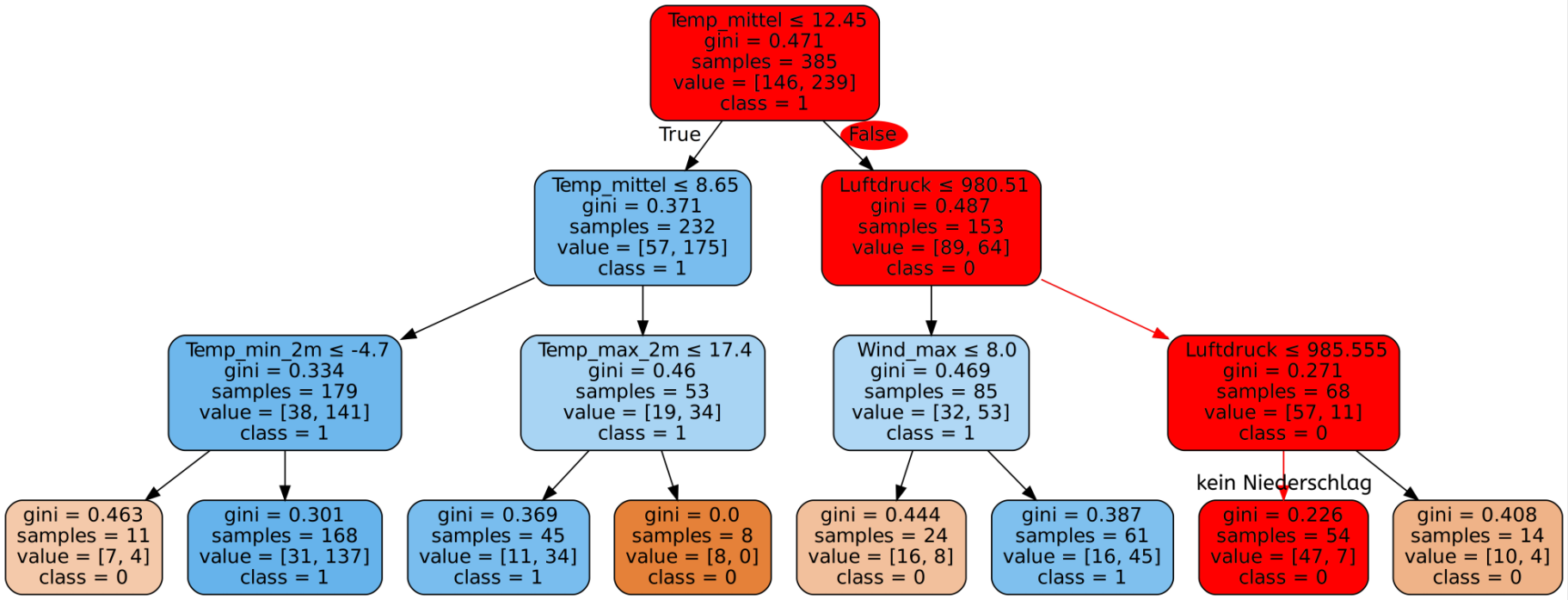
Wind_mittel	relative_Feuchte	Sonnenscheindauer	Bedeckungsgrad	Luftdruck	Temp_mittel	Niederschlagsform
4	57,96	8,117	4,5	978,88	-8,2	1
7	54,33	9,9	2,8	975,46	-7,9	0
6,7	52,21	7,05	2,5	983,71	-6,7	1
2,7	81,25	5	1,7	981,84	-4,9	0
8	48,5	1,633	7,3	964,48	-4,2	1
7	44,46	10,267	1,3	983,96	-4,1	0
1,9	85,63	0	7,3	971,3	-4,1	1
1,2	95,67	0	6,3	993,69	-3,7	1
3,4	79,71	1,633	6,2	976,58	-3,6	0
2,2	71,75	0,117	7,8	971,27	-3,6	1
5,6	68,88	0,05	7,9	957,95	-3,5	1
0,9	94,67	0	6,1	989	-3,3	0
7,1	76,46	0	7,8	965,23	-3,2	1
3,1	83,58	0	8	960,66	-2,6	1
3	77,13	6,883	2,8	976,67	-2,4	0
3,7	81,29	0	8	977,65	-2,1	1
1,5	97,21	0	7,9	991,15	-2	1
6,9	56,13	9,917	0,3	977,81	-1,8	0
3,2	86,29	0	7,8	988,08	-1,8	1
3,5	69,83	2,483	7,4	983,65	-1,8	1
1,7	82,67	7,283	3	974,95	-1,5	0
8,4	51,71	10,067	0,5	979,36	-1,5	0
2,5	87,83	7,95	2,3	987,11	-1,4	0

Testdaten

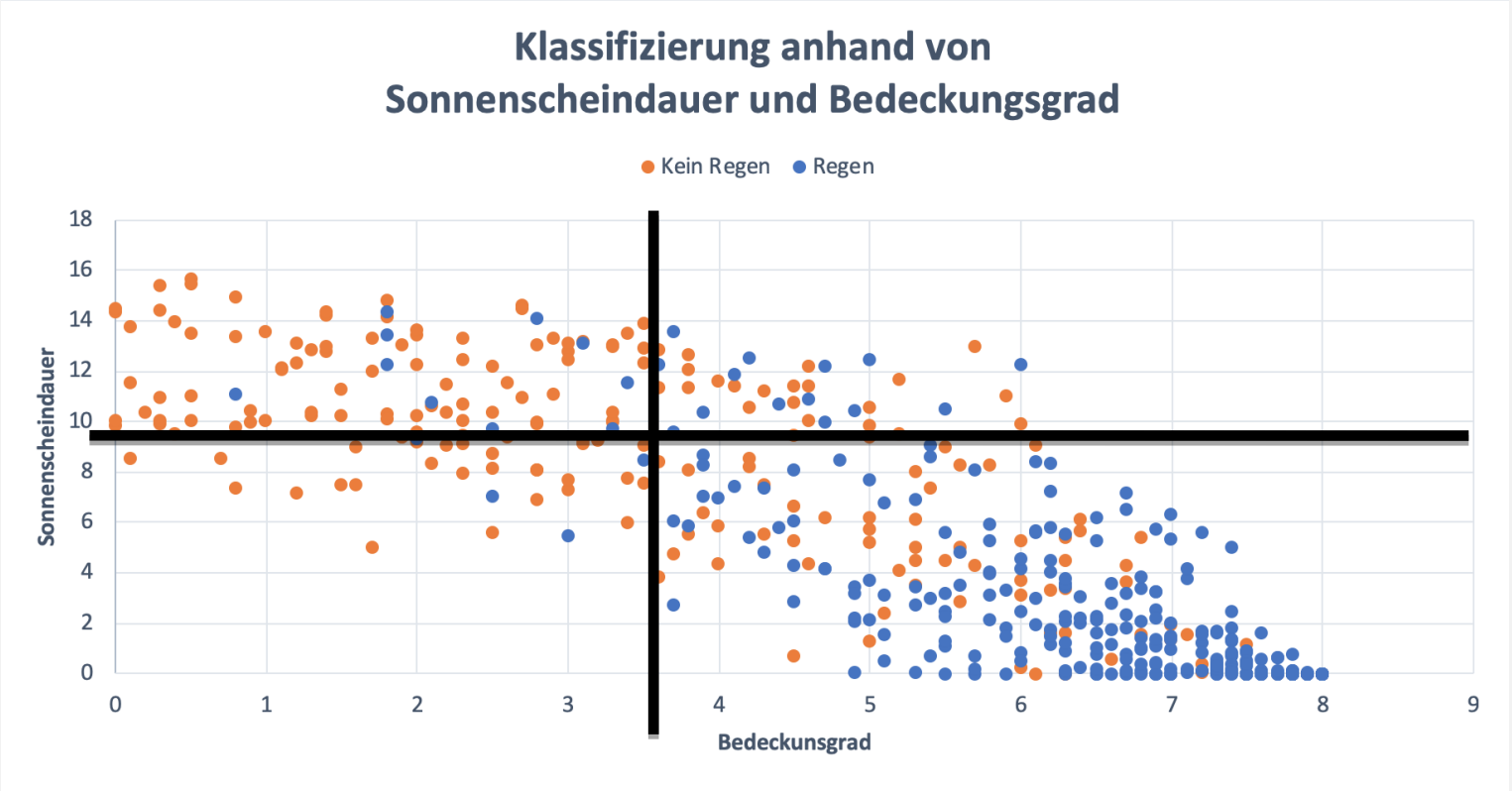
Label

Wie entscheidet der Baum?

Testdaten: Temp_mittel: 20°, Luftdruck 982 hPa, …

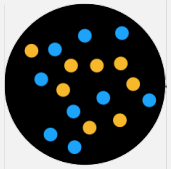


Wie entsteht der Baum?



Wie entsteht der Baum?

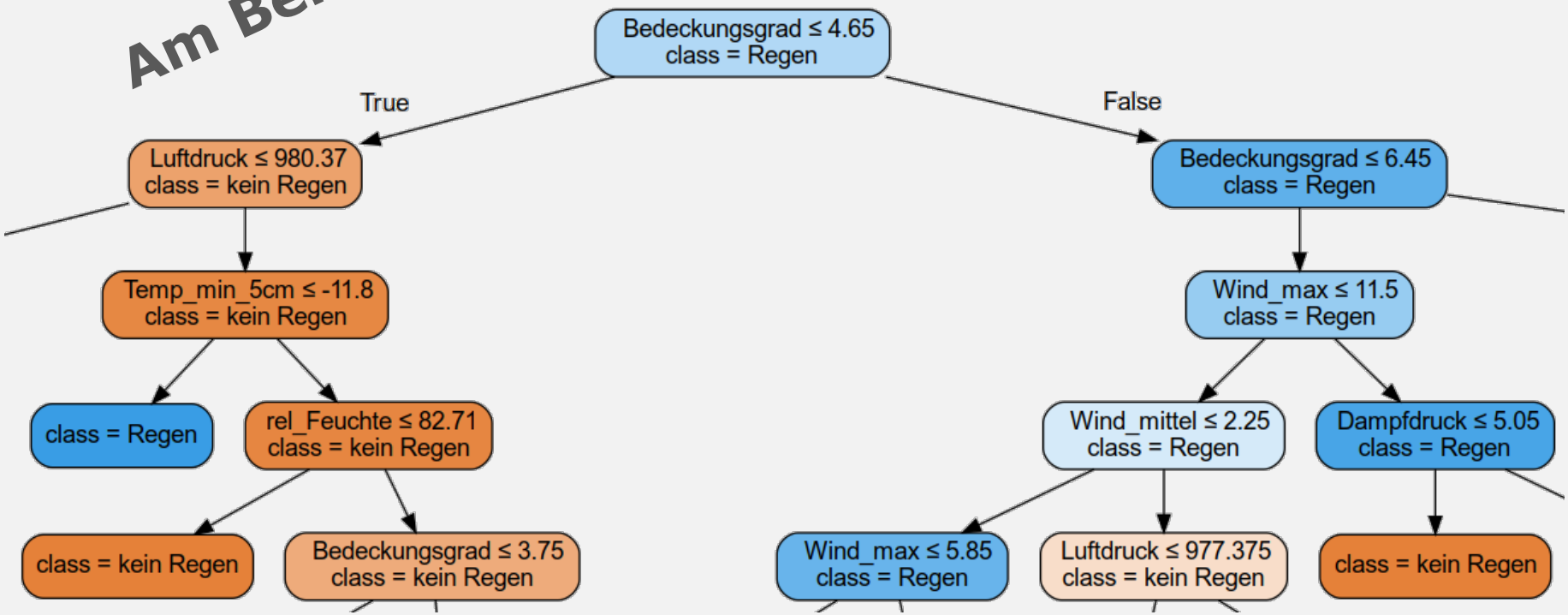
„Bestes Kriterium“ zuerst



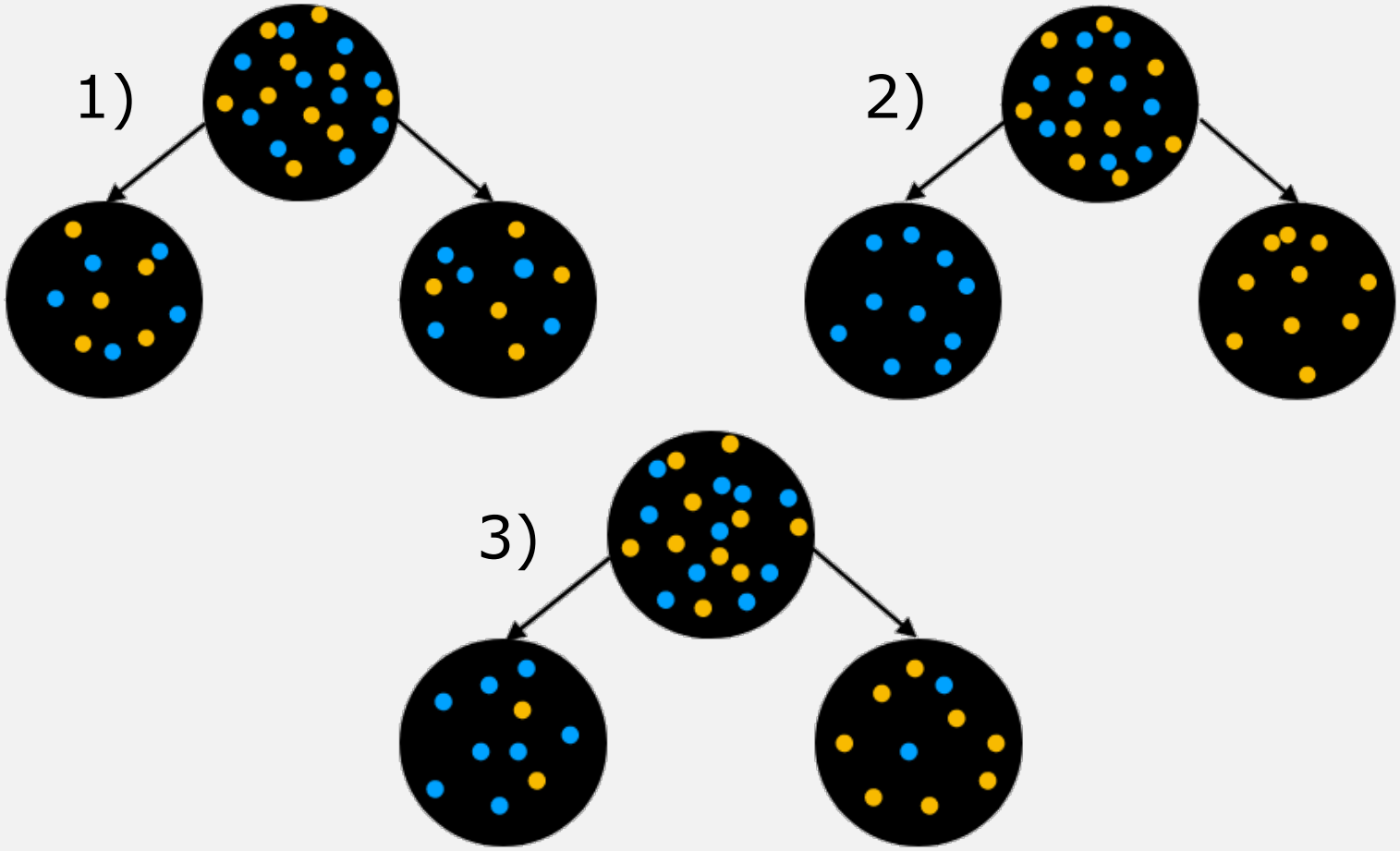
Dann feinere Unterscheidungen

Wie entsteht der Baum?

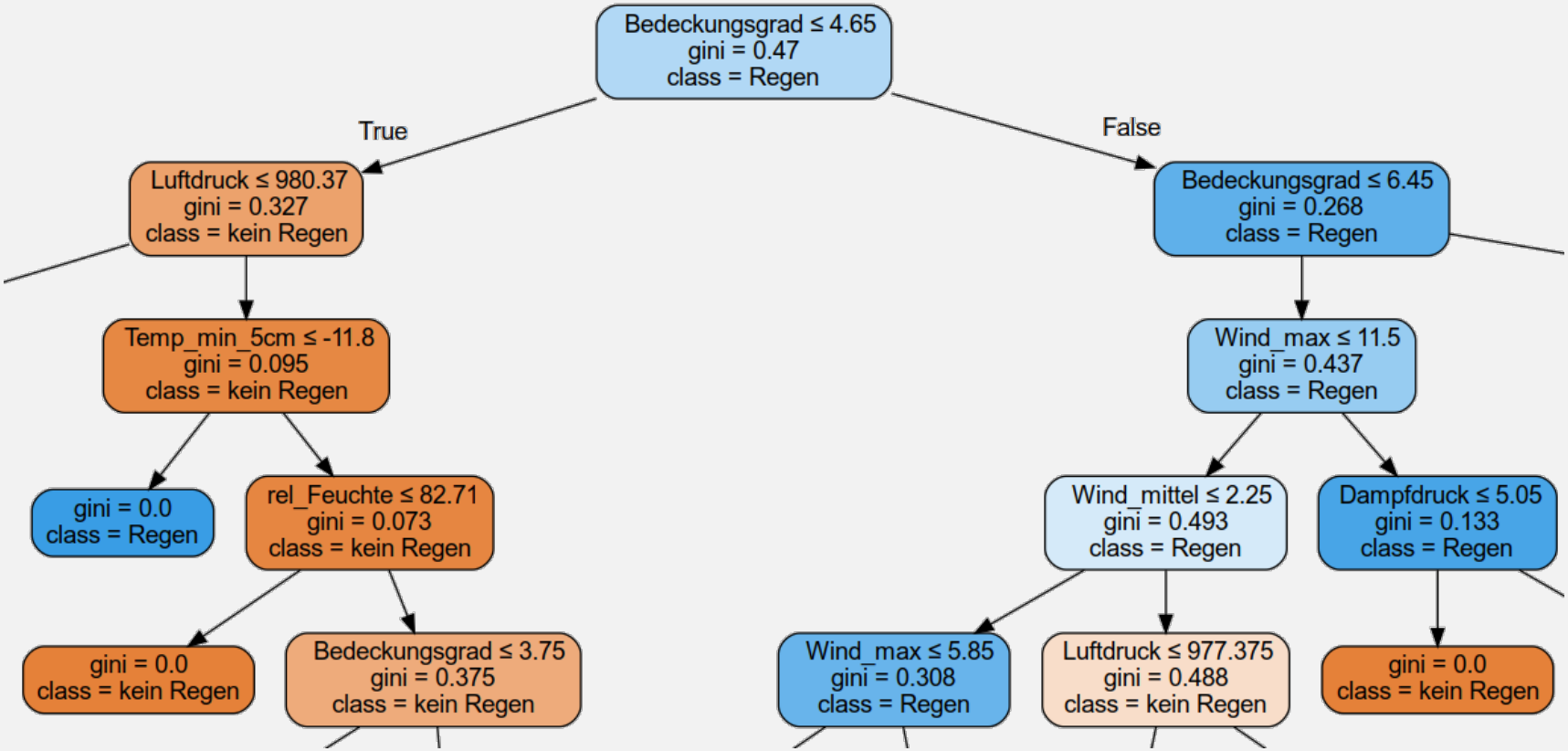
Am Beispiel:



Wie entsteht der Baum?

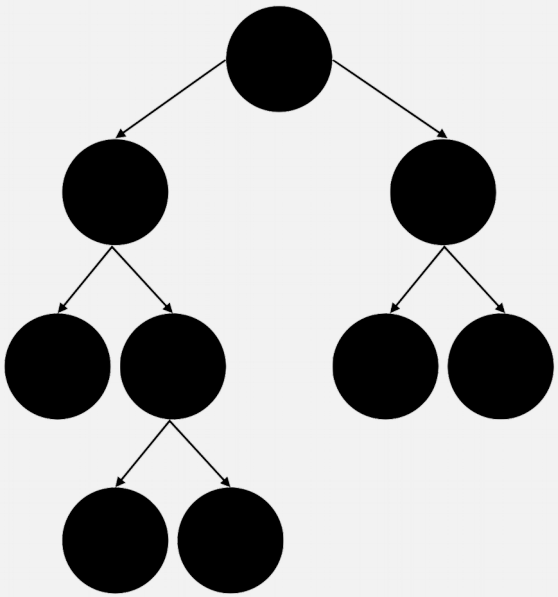


Wie entsteht der Baum?



10 Min Arbeitsphase

Optimierung oder Einsatz zur Klassifizierung



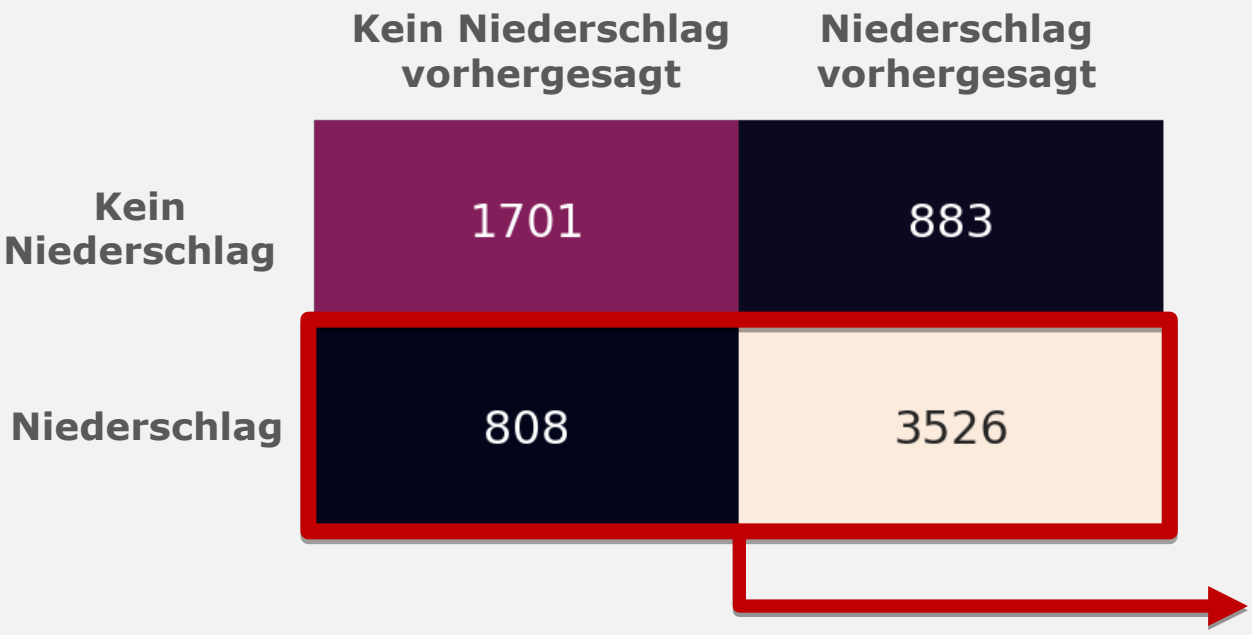
Verbesserung

Ausprobieren
an Testdaten

Arbeitsschritte beim Arbeiten mit ML

1. Daten besorgen
2. Überblick über die Daten verschaffen
3. Aufteilen der Daten in Trainings- und Testdaten
4. Wahl eines Algorithmus
5. Trainieren des Modells
6. Qualität des Modells mithilfe der Testdaten prüfen
7. Optimierung oder Einsatz an neuen Daten

Welche Ergebnisse können auftreten?



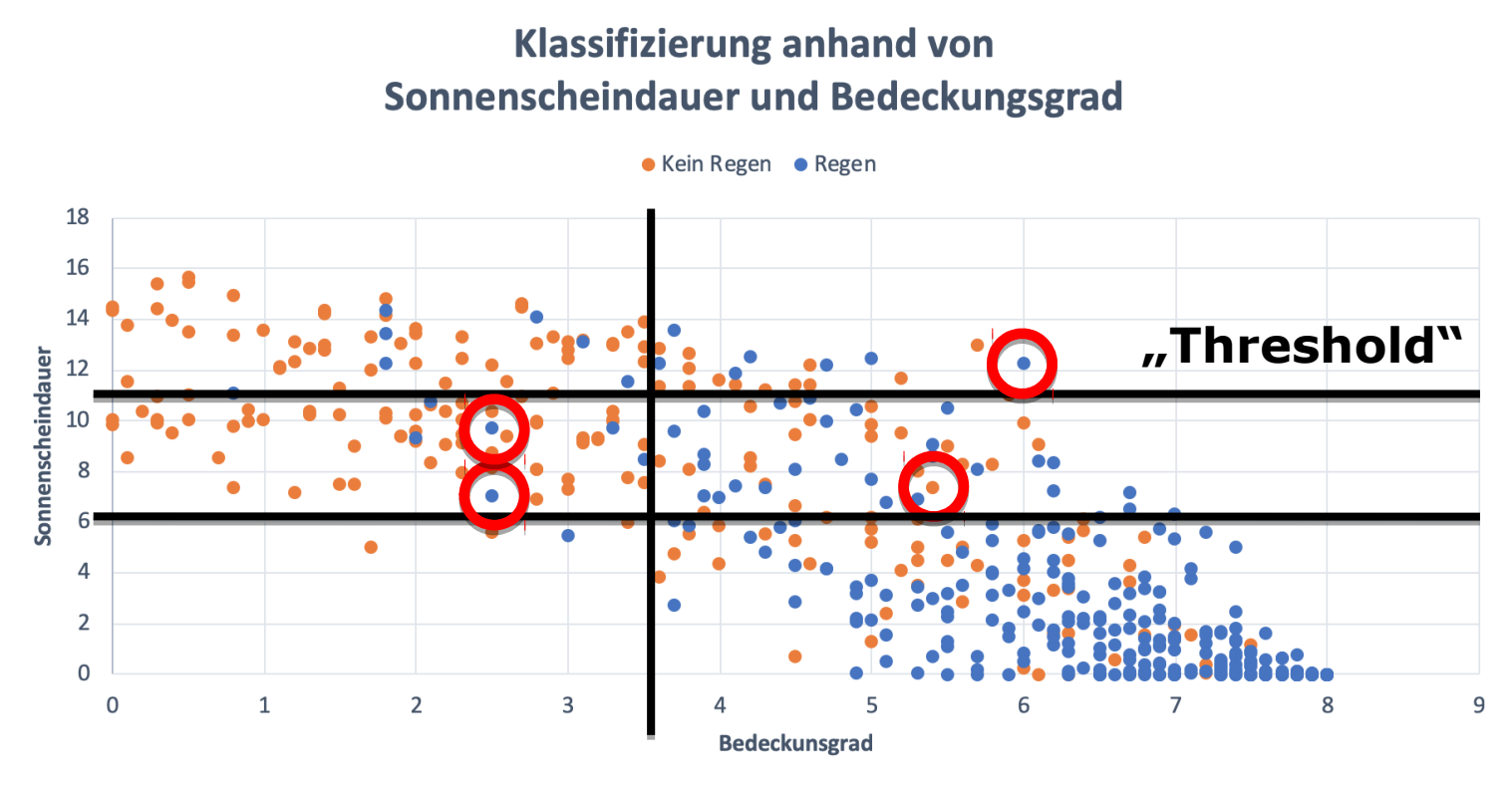
Spezifität:

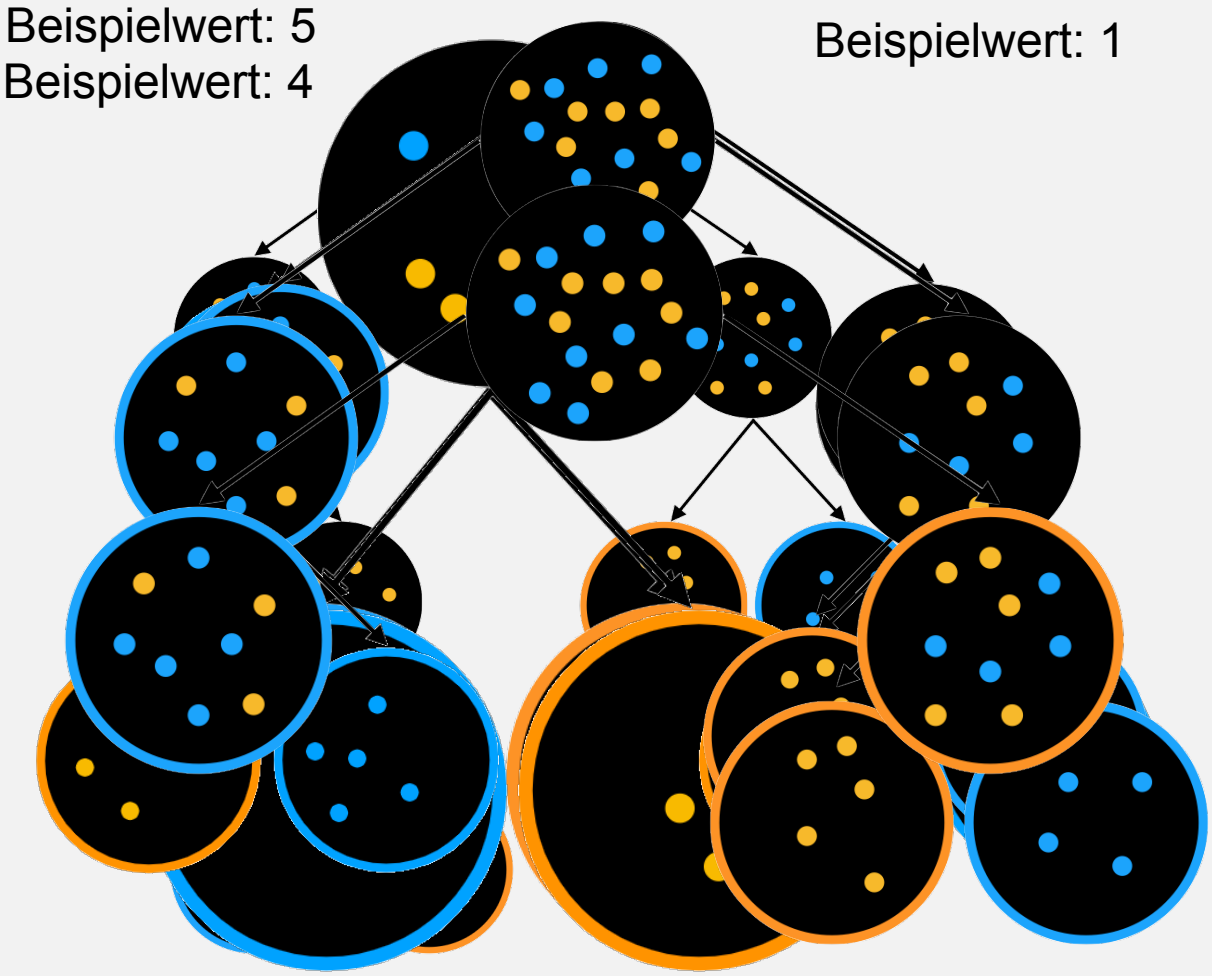
$$\frac{1701}{1701+883}$$

Sensitivität:

$$\frac{3526}{3526+808}$$

Erinnerung


Klassifizierung anhand von
Sonnenscheindauer und Bedeckungsgrad
Kein Regen
Regen
„Threshold“
Sonnenscheindauer
Bedeckunsgrad
0
2
4
6
8
10
12
14
16
18
0
1
2
3
4
5
6
7
8
9



Minimale Anzahl an Datensätzen pro Split

Maximale Anzahl Features pro Split

Minimale Anzahl an Datensätzen pro Blatt

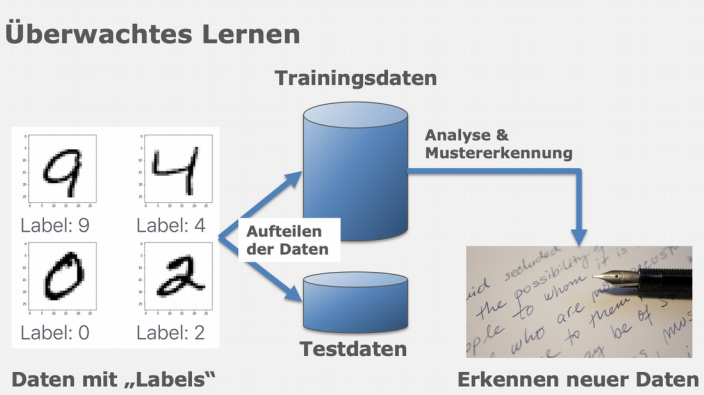
Maximale Tiefe

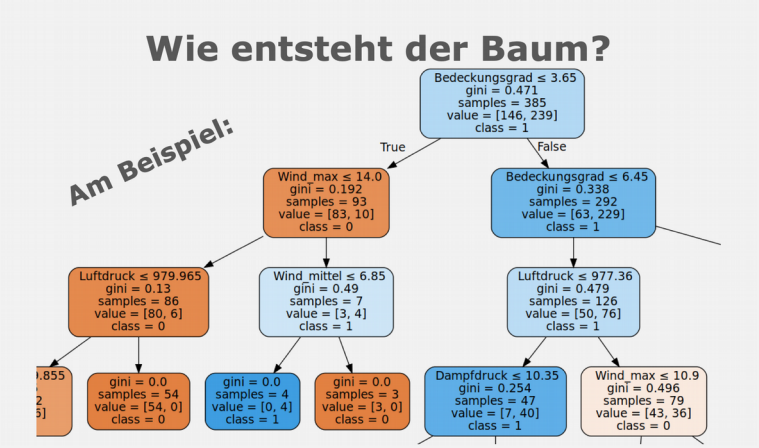
Und jetzt?

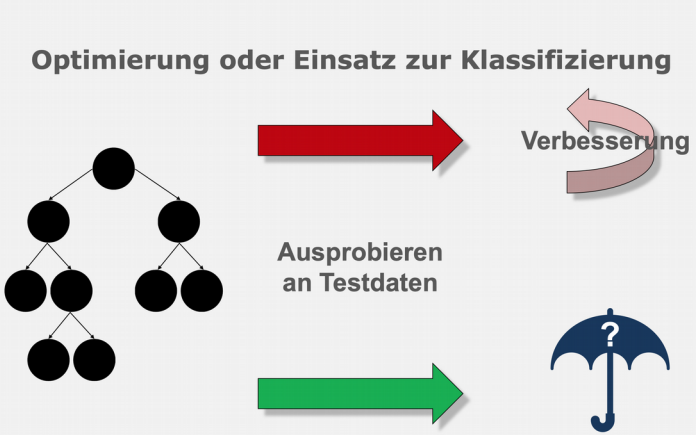
Selbst ausprobieren!

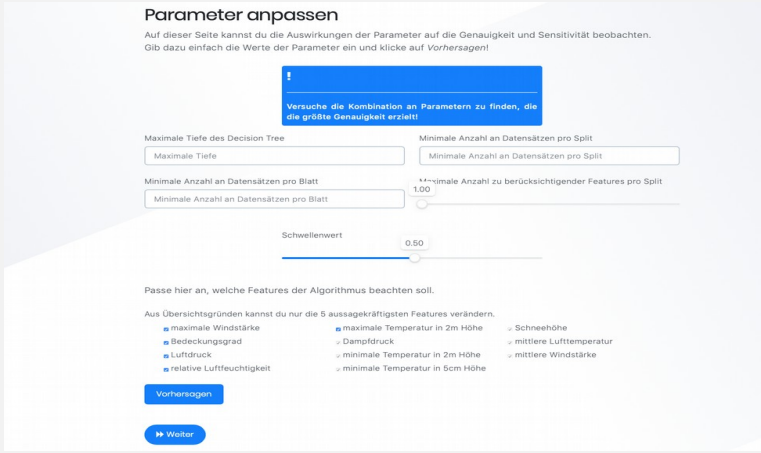
Was konnte beobachtet werden?

Was haben wir gesehen?

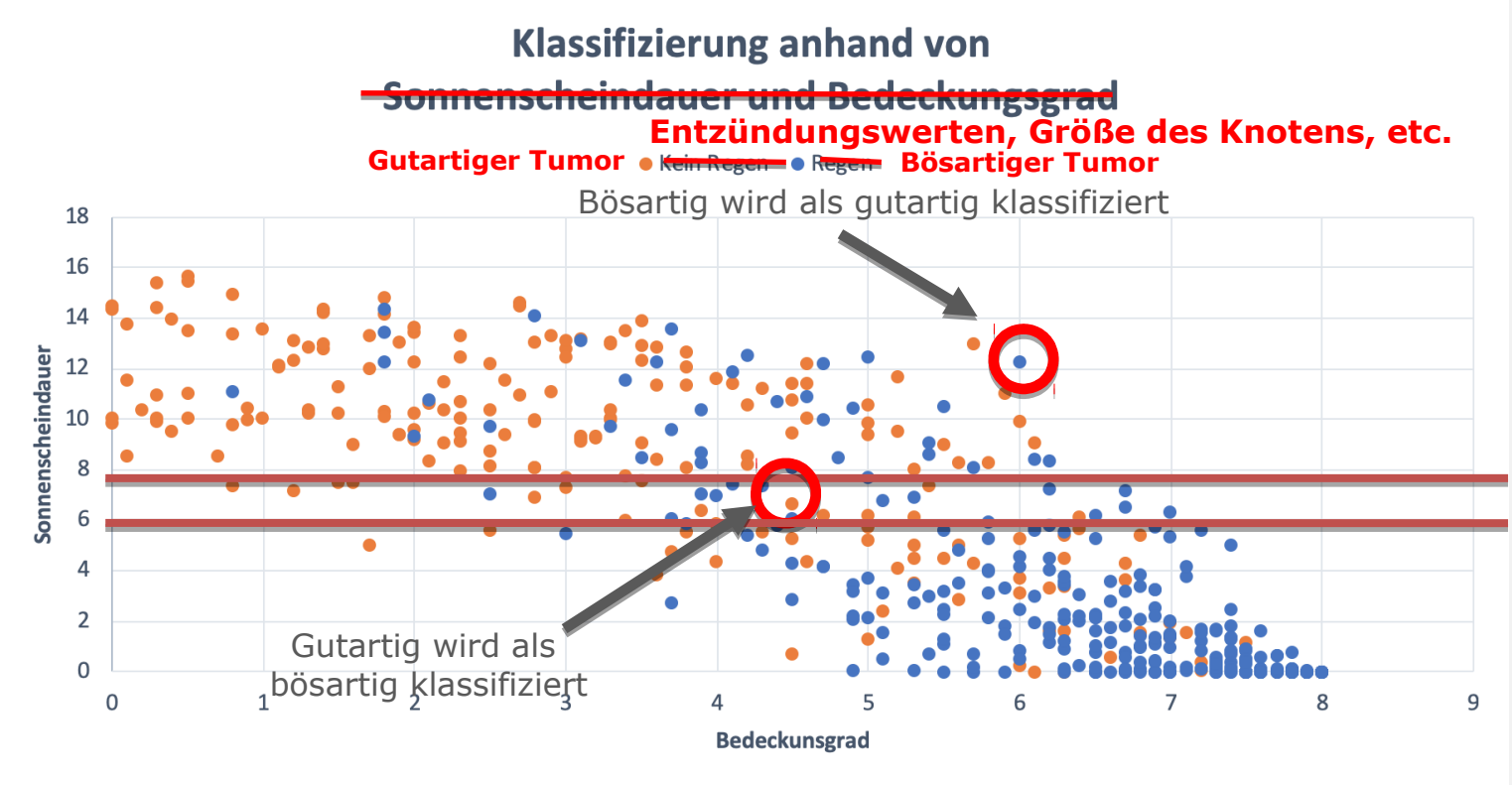








Rekapitulation



Vielen Dank für die Aufmerksamkeit.
© Franz Walgenbach, Lukas Wachter